

АНАЛИЗ И СИНТЕЗ СИГНАЛОВ И ИЗОБРАЖЕНИЙ

УДК 681.3.06

МУЛЬТИАГЕНТНОЕ УПРАВЛЕНИЕ ВЫЧИСЛИТЕЛЬНОЙ СИСТЕМОЙ НА ОСНОВЕ МЕТАМОНИТОРИНГА И ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ*

И. В. Бычков, Г. А. Опарин, А. Г. Феоктистов, И. А. Сидоров,
В. Г. Богданова, С. А. Горский

*Институт динамики систем и теории управления им. В. М. Матросова СО РАН,
664033, г. Иркутск, ул. Лермонтова, 134
E-mail: oparin@icc.ru*

Рассматривается проблема управления вычислениями в распределённой вычислительной среде (РВС) на основе её метамониторинга и имитационного моделирования. Управление вычислениями реализуется мультиагентной системой с заданной организационной структурой. Распределение ресурсов осуществляется агентами с использованием экономических механизмов регулирования их спроса и предложения. Управляющие воздействия на агентов формируются на основе имитационного моделирования процессов функционирования РВС. Сбор информации о ресурсах и процессах РВС, обнаружение и предотвращение нештатных ситуаций в узлах РВС выполняются системой метамониторинга этой среды. Результатами исследований являются методика выбора управляющих воздействий и методы интеллектуальной обработки и эффективного хранения данных.

Ключевые слова: распределённые вычисления, мультиагентное управление, мониторинг.

DOI: 10.15372/AUT20160201

Введение. Анализ тенденций развития технологий высокопроизводительных вычислений как в России, так и за рубежом [1] позволяет сделать вывод о том, что управление параллельными и распределёнными вычислительными системами в настоящее время является одной из важнейших фундаментальных проблем. В качестве вычислительной системы в данной работе рассматривается распределённая вычислительная среда (РВС) с разнородными узлами — вычислительными кластерами (ВК), имеющими сложную гибридную структуру. Гибридный кластер включает в себя вычислительные модули (аппаратные компоненты), поддерживающие различные технологии параллельного программирования и различающиеся своими вычислительными характеристиками. Распределённая вычислительная среда обладает рядом свойств, существенно усложняющих унификацию процессов планирования вычислений и распределения ресурсов, оценки эффективности и надёжности функционирования среды [2]. Можно выделить ряд успешных направлений исследований в области управления распределёнными вычислениями [3–5], в том числе эффективного решения задач с использованием графических процессоров [6]. Однако применение мультиагентных технологий, средств мониторинга и моделирования РВС позволяет в ряде случаев получить более эффективные результаты управления вычислениями и повысить надёжность РВС.

*Работа выполнена при поддержке Российского фонда фундаментальных исследований (проекты № 15-29-07955-офи_м, № 16-07-00931-а).

Целью данного исследования является разработка системы управления вычислениями в РВС, реализующей новые оригинальные мультиагентные методы и средства планирования вычислений и распределения ресурсов на основе метамониторинга и имитационного моделирования РВС, которые, как показывают результаты вычислительного эксперимента, обеспечивают высокую эффективность процессов управления вычислениями и анализа надёжности РВС.

Система управления вычислениями. В настоящее время известны различные агентные методы и средства управления вычислениями [7, 8], используемые на практике, в том числе основанные на экономических принципах распределения вычислительных ресурсов [9–11]. Однако особенности рассматриваемой РВС [2] не позволяют применять их с достаточной эффективностью.

В рамках предлагаемого исследования управление вычислениями реализуется мультиагентной системой (МАС) с заданной организационной структурой. Координация действий агентов осуществляется с помощью правил группового поведения. Агенты функционируют в соответствии с назначенными ролями. Для каждой роли определены свои правила поведения в виртуальном сообществе (ВС) агентов. Мультиагентная система включает в себя агентов метамониторинга и распределения ресурсов, а также управляющего агента. Агенты распределения ресурсов могут объединяться в ВС. В разных ВС, возникающих в МАС, агенты координируют свои действия путём кооперации или соперничества.

Целью МАС является получение распределения поступающих в РВС потоков заданий, сохраняющего показатели качества функционирования РВС в пределах, установленных её администратором. Задание представляет собой спецификацию процесса решения задач, содержащую информацию о требуемых вычислительных ресурсах, исполняемых прикладных программах, входных/выходных данных, а также другие необходимые сведения.

Структурная схема системы управления вычислениями представлена на рис. 1. На схеме РВС выступает в качестве объекта управления, для которого потоки w_1 и w_2 заданий пользователей РВС и ВК являются внешними возмущениями, результаты распределения d_1 и d_2 потоков w_1 и w_2 по ВК — управляющими воздействиями МАС и пользователей ВК соответственно, а вектор r_1 параметров административных политик ВК — корректирующим воздействием. Агенты распределения ресурсов перехватывают задания потока w_1 в целях более детальной настройки требований к РВС, содержащихся в заданиях, при этом поток w_1 модифицируется в поток w'_1 . Распределение d_1 потока w'_1 производится агента-

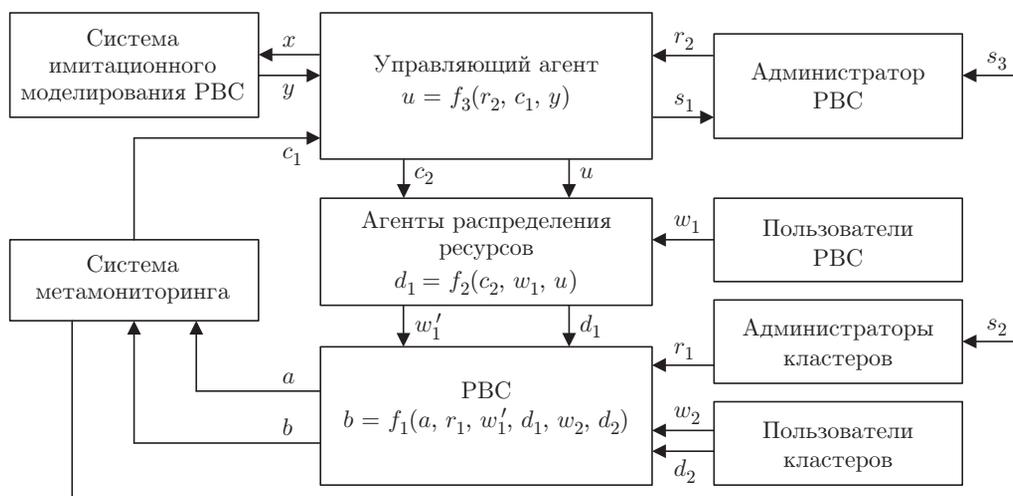


Рис. 1

ми на основе экономических механизмов [12]. Распределение d_2 потока w_2 определяется пользователями ВК.

Сведения о вычислительных характеристиках узлов РВС и текущих показателях объёмов работ в этих узлах собираются агентами системы метамониторинга [13] с помощью контрольно-измерительных средств в виде структур данных a и b . Между компонентами структуры b , с одной стороны, и элементами структуры a , корректирующим воздействием r_1 , потоками заданий w'_1 и w_2 , распределениями d_1 и d_2 , с другой стороны, существует некоторая абстрактная связь $b = f_1(a, r_1, w'_1, d_1, w_2, d_2)$. Для различных компонентов структуры b эта связь представлена функциональным, статистическим, неоднозначным или иным отображением. Связи f_2 и f_3 имеют ту же природу, что и связь f_1 .

Собранные сведения в виде вектора c_1 агрегированных показателей функционирования объекта управления передаются управляющему агенту по его запросу, посылаемому с некоторым периодом дискретности T_q . Величина T_q выбирается так, чтобы не перегружать коммуникационную сеть РВС и в то же время с необходимой точностью фиксировать моменты приближения показателей функционирования объекта управления к их предельным значениям. Часть сведений, представленных вектором c_1 и актуальных для агентов распределения ресурсов, немедленно передаётся этим агентам в виде вектора c_2 .

Вектор r_2 параметров административных политик РВС является корректирующим воздействием для управляющего агента. На основе информации, представленной векторами c_1 и r_2 , в установленные моменты времени управляющий агент прогнозирует на определённый промежуток времени динамику показателей качества функционирования объекта управления с помощью системы имитационного моделирования с периодом дискретности $T_s > T_q$. Результаты моделирования используются для формирования вектора управляющих воздействий u на алгоритмы работы агентов распределения ресурсов ВК путём параметрической настройки алгоритмов. После того как вектор u сформирован, он передаётся каждому ВК.

Имитационное моделирование РВС. Пусть x и y — векторы входных и наблюдаемых переменных имитационной модели РВС. Наблюдаемые переменные представляют показатели качества функционирования РВС. Элементы векторов x_i , $i = \overline{1, n_x}$, и y_j , $j = \overline{1, n_y}$, имеют соответственно области X_i и Y_j допустимых значений. Эффекты влияния входных переменных на наблюдаемые исследованы с помощью факторного анализа заранее: при построении и испытании имитационной модели РВС. Также предполагается, что для каждого j -го элемента вектора y заданы критерий вычисления оценки \hat{y}_j качества значения этого элемента (стремление к минимуму или максимуму) и его границы y_j^{\min} , $y_j^{\max} \in Y_j$. Ряд элементов вектора x играют роль варьируемых переменных, образуют подмножество X^* и отождествляются с элементами вектора u : $u_q \equiv x_i$, $q = \overline{1, n_u}$, $i \in \overline{1, n_x}$, $1 \leq n_u < n_x$. В качестве начальных значений варьируемых переменных используются базовые, соответствующие принятым по умолчанию параметрам функционирования РВС. Последующие значения выбираются из их доменов с учётом эффектов влияния x_i на y_j , $i \in \overline{1, n_x}$, $j = \overline{1, n_y}$. Значения неварьируемых переменных, являющихся элементами вектора x , задаются на основе числовой информации, представленной векторами r_2 и c_1 .

В процессе моделирования формируется множество V вариантов значений наблюдаемых переменных: величина $y_{jk} \in Y_j$ есть элемент k -го варианта $v_k \in V$ для переменной y_j , $j = \overline{1, n_y}$, $k = \overline{1, n_v}$. Формирование подмножества $V^* \subseteq V$ вариантов значений наблюдаемых переменных на основе множества V в целях дальнейшего определения элементов вектора u является многокритериальным. Выбор вариантов для V^* осуществляется на основе лексикографического метода, если наблюдаемые переменные упорядочены по значимости, в противном случае — на основе мажоритарного метода. Лексикографический и

мажоритарный методы используют правила выбора [14]

$$V^* = \{v_k \in V: (\forall v_l \in V \exists p \in \overline{1, n_y - 1}: (\hat{y}_{1k} = \hat{y}_{1l}) \wedge \dots \wedge (\hat{y}_{pk} = \hat{y}_{pl}) \wedge (\hat{y}_{(p+1)k} > \hat{y}_{(p+1)l}))\}, \quad (1)$$

где $y_j^{\min} \leq y_{jk} \leq y_j^{\max}$, $j \in \overline{1, n_y}$; $k \in \overline{1, n_v}$; $l \in \overline{1, n_v}$; $k \neq l$,

$$V^* = \left\{ v_k \in V: \left(\neg \exists v_l \in V: \sum_{j=1}^{n_y} \text{sign}(\hat{y}_{jl} - \hat{y}_{jk}) > 0 \right) \right\}, \quad (2)$$

где $\text{sign}(0) = 0$; $y_j^{\min} \leq y_{jk} \leq y_j^{\max}$; $k \in \overline{1, n_v}$; $l \in \overline{1, n_v}$; $k \neq l$.

Для оценки значений y_{jk} , $k \in \overline{1, n_v}$, j -й переменной их множество разбивается на подмножества, которые попарно не пересекаются и упорядочиваются по возрастанию или убыванию. В соответствии с этим каждое подмножество получает свой индекс, применяющийся в качестве оценки значений наблюдаемых переменных, попавших в данное подмножество.

Пусть в V^* включён единственный вариант v_k , с которым согласуется k -й набор варьируемых переменных вектора x . Выбрав из них x_{ik} (при $x_i \in X^*$), получим значения элементов вектора u . Если число вариантов в V^* больше одного, то выбор единственного v_k осуществляется случайным образом. При $V^* = \emptyset$ управляющий агент порождает сигнал s_1 , требующий от администратора РВС новых корректирующих воздействий.

Система метамониторинга РВС. Узлы РВС, как и любые другие технические системы, подвержены отказам и сбоям их компонентов. При большом числе узлов РВС такие нештатные ситуации приводят к прекращению выполнения параллельных заданий и необходимости их рестарта, как правило, с помощью контрольных точек. Рестарты сопровождаются дополнительными накладными расходами и влекут снижение эффективности функционирования РВС [15].

Система метамониторинга РВС предназначена для обнаружения и предотвращения нештатных ситуаций в узлах. С некоторым периодом дискретности T_i агенты метамониторинга измеряют значение характеристики h_i узла РВС с помощью локальных систем мониторинга m_j (например, Ganglia и Nagios [16]) и контрольно-измерительных средств e_k (например, Sensors, APC PowerChute, IPMI и SMART [17]), $i \in \overline{1, n_h}$, $j \in \overline{1, n_m}$, $k \in \overline{1, n_e}$. Для каждой i -й характеристики заданы её предельные значения h_i^{\min} и h_i^{\max} : $h_i^{\min} \leq h_i \leq h_i^{\max}$. Всё множество характеристик узла делится на три подмножества:

1) характеристики объёмов вычислительной нагрузки компонентов узла — процессоров, ядер, оперативной памяти, сетевых элементов, систем хранения данных и других структурных элементов;

2) характеристики физического состояния компонентов узла — температура процессоров и материнских плат, работоспособность систем бесперебойного питания, жёстких дисков и других структурных элементов;

3) характеристики процесса выполнения задания — приоритет и статус задания, процессорное время, объём оперативной памяти, число обращений к жёсткому диску и сетевым элементам и другая информация.

Вследствие большого объёма сведений о характеристиках узлов РВС в системе метамониторинга реализована оригинальная циклическая база данных [18], показавшая более высокую эффективность процессов накопления, унификации и агрегирования информации в сравнении с аналогичными известными базами, например MRTG [19] или RRDtool [20].

Система управления	Показатели эффективности системы управления вычислениями					
	n_{avg} , единицы	t_{avg} , с	k_{avg} , %	σ	n_{rest} , единицы	n_{err} , единицы
GridWay	520,362	701,750	0,690	0,005	123	31
MAC	98,071	286,184	0,746	0,003	57	1

На основе собранных сведений в системе метамониторинга выполняется экспертный анализ характеристик узлов РВС. В случае приближения значения характеристики h_i узла к её границе h_i^{\min} или h_i^{\max} система метамониторинга порождает сигналы s_2 и s_3 , требующие от администратора кластера, в который входит узел, и администратора РВС новых корректирующих воздействий. Анализ характеристик процессов выполнения заданий в узлах позволяет выявлять в том числе неэффективную работу пользовательских приложений и учитывать эти результаты агентами распределения ресурсов. Так, к примеру, для программного пакета Maker [21] было обнаружено преобладание операций записи в сетевую директорию по отношению к вычислительным операциям, что связано с низкой производительностью системы хранения информации в узлах. Запуск заданий в узлах с подключённым локальным хранилищем способствовал росту эффективности выполнения пакета программ Maker более чем на 30 %. Экспертная подсистема реализована в среде CLIPS [22]. Периоды дискретности измерения характеристик и их экспертный анализ обеспечивают оперативное реагирование системы метамониторинга на нештатные ситуации с аппаратными средствами РВС и тем самым повышают надёжность и эффективность функционирования РВС.

Результаты вычислительного эксперимента. В целях более полного исследования рассмотренных выше методов и инструментальных средств проведено имитационное моделирование функционирования РВС с помощью системы GPSS World [23]. Моделируемая система состояла из 10 кластеров (с числом ядер от 6000 до 14000 единиц) и 300 пользователей. Общее число ядер составляло 100000 единиц. Кластеры включали в себя гибридные узлы, поддерживающие различные технологии параллельного программирования. При имитации времени выполнения задания использовались коэффициенты ускорения счёта, значения которых для разных кластеров варьировались от 1 до 1,5 в зависимости от их вычислительных характеристик. Моделируемый период времени работы системы — 30 суток. За этот период было обработано 12990 потоков заданий, включавших от 1000 до

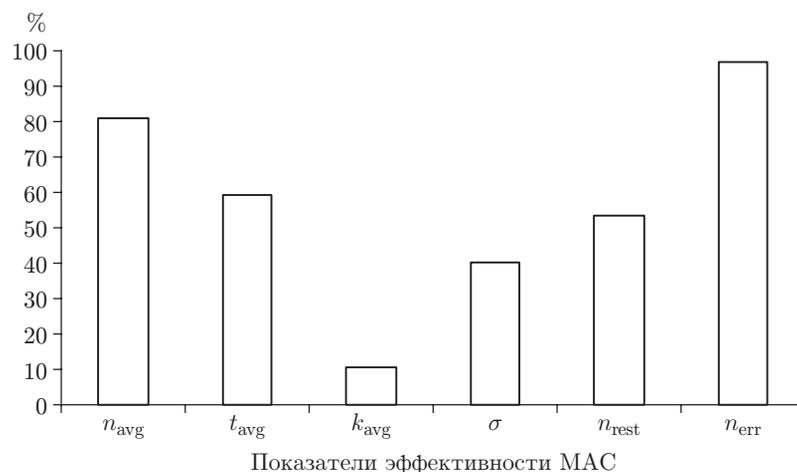


Рис. 2

10000 процессов для параллельных программ или многовариантных расчётов. В качестве систем управления вычислениями в РВС применялись метапланировщик GridWay [24] и МАС, представленная в данной работе. Дисциплина обслуживания очередей заданий — FCFS (First Come, First Served) с приоритетами. В качестве основных наблюдаемых переменных имитационной модели были выбраны следующие средние показатели: n_{avg} — число заданий в очереди кластера; t_{avg} — время пребывания задания в очереди кластера; k_{avg} — коэффициент полезного использования узлов кластеров; n_{rest} — число рестартов программ; n_{err} — число сбойных задач; σ — среднеквадратическое отклонение коэффициента полезного использования узлов кластеров. Приведённые в таблице результаты моделирования показывают, что применение МАС может существенно улучшить все выбранные показатели функционирования РВС по сравнению с метапланировщиком GridWay (рис. 2).

Заключение. В данной работе исследована проблема управления вычислениями в РВС на основе её метамониторинга и имитационного моделирования. Создана оригинальная мультиагентная система управления вычислениями. Предложена методика многокритериального выбора управляющих воздействий МАС. Для повышения эффективности и надёжности функционирования РВС реализованы методы децентрализованной интеллектуальной обработки и распределённого хранения данных. Проведён модельный вычислительный эксперимент. Показано, что созданная МАС превосходит широко используемый на практике метапланировщик GridWay по целому ряду важных показателей эффективности управления вычислениями.

СПИСОК ЛИТЕРАТУРЫ

1. **Шамакина А. В.** Обзор технологий распределённых вычислений // Вестн. ЮУрГУ. Сер. Вычислительная математика и информатика. 2014. **3**, № 3. С. 51–85.
2. **Bogdanova V. G., Bychkov I. V., Korsukov A. S. et al.** Multiagent approach to controlling distributed computing in a cluster grid system // Journ. Comput. Syst. Sci. Intern. 2014. **53**, N 5. P. 713–722.
3. **Коваленко В. Н., Коваленко Е. И., Корягин Д. А. и др.** Основные положения метода опережающего планирования для грид вычислительного типа // Вестн. СамГУ. Естественно-научная сер. 2006. № 4. С. 238–264.
4. **Kononov M. G., Malashenko Yu. E., Nazarova I. A.** Job control in heterogeneous computing systems // Journ. Comput. Syst. Sci. Intern. 2011. **50**, N 2. P. 220–237.
5. **Toporkov V. V.** Job control in distributed environments with non-dedicated resources // Journ. Comput. Syst. Sci. Intern. 2011. **50**, N 3. P. 413–428.
6. **Якименко А. А., Гунбин К. В., Хайретдинов М. С.** Поиск перепредставленных характеристик генов: опыт реализации перестановочного теста с использованием графических процессоров // Автометрия. 2014. **50**, № 1. С. 123–129.
7. **Durfee E. H.** Distributed problem solving and planning // Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence /Ed. G. Weiss. Cambridge: MIT Press, 1999. P. 121–164.
8. **Altameem T., Amoon M.** An agent-based approach for dynamic adjustment of scheduled jobs in computational grids // Journ. Comput. Syst. Sci. Intern. 2010. **49**, N 5. P. 765–772.
9. **Mutz A., Wolski R., Brevik J.** Eliciting honest value information in a batch-queue environment // Proc. of the 8th IEEE/ACM Intern. Conf. on Grid Computing. IEEE, 2007. P. 291–297.
10. **Market-Oriented Grid and Utility Computing** /Eds. R. Buyya, K. Bubendorfer. Hoboken: Wiley & Sons, 2010. 643 p.

11. **Toporkov V. V., Yemelyanov D. M.** Economic model of scheduling and fair resource sharing in distributed computations // Programming and Computer Software. 2014. **40**, N 1. P. 35–42.
12. **Бычков И. В., Опарин Г. А., Феоктистов А. Г. и др.** Мультиагентный алгоритм распределения вычислительных ресурсов на основе экономического механизма регулирования их спроса и предложения // Вестн. компьютерных и информационных технологий. 2014. № 1. С. 39–45.
13. **Опарин Г. А., Новопашин А. П., Сидоров И. А. и др.** Система метамониторинга распределительных вычислительных сред // Программные продукты и системы. 2014. № 2. С. 45–48.
14. **Шоломов Л. А.** Логические методы исследования дискретных моделей выбора. М.: Наука, 1989. 288 с.
15. **Бетелин В. Б., Кушниренко А. Г., Райко Г. О.** Проблемы обеспечения роста производительности отечественных суперЭВМ в период до 2020 года // Информационные технологии и вычислительные системы. 2010. № 3. С. 15–18.
16. **Zanikolas S., Sakellariou R.** A taxonomy of grid monitoring services // Future Generat. Comput. Syst. 2005. **21**, N 1. P. 163–188.
17. **Charoenpornwattana K.** A Scalable Unified Fault Tolerance for High Performance Computing Environments. Ruston, USA: Louisiana Tech University, 2008. 132 p.
18. **Сидоров И. А., Новопашин А. П., Опарин Г. А. и др.** Методы и средства метамониторинга распределительных вычислительных сред // Вестн. ЮУрГУ. Сер. Вычислительная математика и информатика. 2014. **3**, № 2. С. 30–42.
19. **MRTG** — The Multi Router Traffic Grapher. URL: <http://oss.oetiker.ch/mrtg/> (дата обращения: 04.08.2015).
20. **RRDTool**. URL: <http://www.rrdtool.org/> (дата обращения: 04.08.2015).
21. **MAKER** — Genome Annotation Pipeline. URL: <http://gmod.org/wiki/MAKER> (дата обращения: 04.08.2015).
22. **Частиков А. П., Белов Д. Л., Гаврилова Т. А.** Разработка экспертных систем. Среда CLIPS. С.-Пб.: БХВ-Петербург, 2003. 393 с.
23. **Боев В. Д.** Моделирование систем. Инструментальные средства GPSS World. С.-Пб.: БХВ-Петербург, 2004. 368 с.
24. **Herrera J., Huedo E., Montero R. et al.** Porting of scientific applications to grid computing on GridWay // Scientific Programming. 2005. **13**, N 4. P. 317–331.

Поступила в редакцию 3 сентября 2015 г.
